Using Ridge Regression in Systematic Pointing Error Corrections

C. N. Guiar
Ground Antenna and Facilities Engineering Section

A pointing error model is used in the antenna calibration process. Data from spacecraft or radio star observations are used to determine the parameters in the model. However, the regression variables are not truly independent, displaying a condition known as multicollinearity. Ridge regression, a biased estimation technique, is used to combat the multicollinearity problem. Two data sets pertaining to Voyager 1 spacecraft tracking (days 105 and 106 of 1987) were analyzed using both linear least squares and ridge regression methods. The advantages and limitations of employing the technique are presented. The problem is not yet fully resolved.

I. Introduction

A pointing error model is used in the antenna calibration process to compensate for systematic error sources. Data from spacecraft (s/c) or radio star observations are used to determine the parameters in the model. The model parameters are then used to generate a systematic error correction table for accurately pointing the antenna. The pointing error modeling approach used was originally devised by optical astronomers and subsequently adapted by radio astronomers for RF antennas. The model is based on logical, expected physical behavior of the antenna and has been successfully applied to many radio astronomy facilities: the Bonn 100-m Az-El antenna [1] and the Haystack 37-m Az-El antenna [2]. The complete pointing error model for an antenna is a sum of individual error functions. Table 1 shows the individual error sources and the elevation and cross-elevation (or, depending on the antenna mount, declination and cross-declination) error functions used to develop a systematic error correction table ([1], [2] and [3] give a more in-depth description of the parameters).

When modeling a system, one may select the model purpose to fall into one of three main categories: explanation, variable selection, or prediction. If the model is explanatory, then it represents the y in terms of the x's and explains how the x's affect the y. Variable selection techniques should be used when the goal is to determine which variables from a group of variables are important in determining the optimal model for y. This selection of variables could provide the best fit, the simplest form of the model, or both. Prediction, or forecasting, techniques estimate the output, y, at previously unobserved values of inputs, x.

The current pointing error model used in the DSN is of the explanatory type, and the parameters P are determined by performing a linear least squares fit on offset data collected from s/c or radio star observations. Currently, the regressor variables are not truly independent and, rather, display redundant information—a condition known as multicollinearity [4]. Multicollinearity results in limitations on the ability of an ordinary linear least squares fit to provide stable and accurate variables.

It is therefore desirable to study alternate techniques for parameter estimation. Ridge regression is a biased estimation technique for combating the multicollinearity problem. This article reviews the use of the ridge regression technique and demonstrates the advantages and limitations of its uses for systematic error correction development.

II. Review of Regression Analysis

Suppose that, in an experiment, values of the dependent variable y are observed, each corresponding to a particular value of an independent variable x. A straight line representation of the $y = y(x)$ data would have the form

$$y = \beta_0 + \beta_1 x + \epsilon \tag{1}$$

where ϵ is the model error. Equation (1) is a simple linear regression model since it contains a single regressor variable, x, and is linear in x.

The above linear regression of y upon a single variable x can be extended to the multiple linear regression model

$$y_i = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + \ldots + \beta k x_{ki} + \epsilon_i \tag{2}$$

where $i = 1, 2, \ldots, n$ $(n \geqslant k + 1)$, ϵ_i is a conceptual random model error assumed to be uncorrelated for each observation (having a zero mean and a constant variance σ^2), x_{ki} are the independent variables (or regressors), y_i are the dependent variables (or response variables) and are the true responses, and β_k are the unknown regression parameters. One equation can be written for each observation, and the error term ϵ allows the model to be an equality. In matrix terms, Eq. (2) becomes

$$\underline{y} = \underline{\beta}\mathbf{X} + \underline{\epsilon} \tag{3}$$

Since the regression terms β_k are unknown, let the least squares estimator for these coefficients be b_k. These estimators should satisfy the following equation:

$$\hat{y}_i = b_0 + b_1 x_{i1} + \ldots + b_k x_{ik} \tag{4}$$

where $\hat{y}_i$ are the model's estimated (or fitted) value to y_i of Eq. (2). Since Eq. (4) contains only known terms, it does not contain the conceptual terms ϵ_i.

If the initial model was accurate, then the difference between y_i and $\hat{y}_i$ should be small. The difference or residual, r_i, between the actual values and the fitted values is

$$r_i = y_i - \hat{y}_i \tag{5}$$

The method of least squares chooses b_{ik} values so that

$$\sum_{i=1}^{n} r_i^2$$

is minimized. The estimates satisfy the following matrix equation [4], [5]:

$$\underline{b} = (\mathbf{X}'\mathbf{X})^{-1}\,\mathbf{X}\underline{y} \tag{6}$$

where $\mathbf{X}'$ is the transpose of $\mathbf{X}$. When the regressor variables are centered (made dimensionless relative to a mean value), $\mathbf{X}'\mathbf{X}$ is then in correlation form and will be written as $\mathbf{X}^{*\prime}\mathbf{X}^{*}$.

III. Multicollinearity

Multicollinearity exists when the regressor variables are empirically correlated, affecting the computation of $\underline{b}$, which involves the $\mathbf{X}'\mathbf{X}$ matrix. When this situation exists, no conclusions can be drawn as to the individual roles of the variables. If multicollinearity is "severe," then the coefficients may (1) be the wrong size (too large in magnitude); (2) have the wrong sign; or (3) be unstable due to ill-conditioned matrix computations (i.e., small changes in the y's or x's lead to large changes in the coefficients). Multicollinearity will also inhibit the ability to predict.

Diagnostics can be performed to evaluate the extent of the multicollinearity problem. Large values in the correlation matrix are one indication of multicollinearity, but this observation only shows pairwise correlations, not correlations that exist between more than two variables. Variance Inflation Factors (VIFs) are another means of identifying multicollinearity. VIFs are the diagonal elements of the inverse of the correlation matrix and represent the inflation that each regression coefficient experiences above the ideal (identity matrix). VIFs are considerably more useful for multicollinearity detection than simple correlation values because they give a direct measure of multicollinearity and tell the user which coefficients are adversely affected and to what extent. As a rule of thumb, VIFs greater than 10 indicate that a severe multicollinearity problem exists. Table 2 gives a sample analysis of a set of conical scanning (conscan) offset data (collected during a Voyager 1 track on the 105th day of 1987) that exhibits a multicollinearity problem. Correlation values of zero mean no correlation and ±1.0 means full correlation. The VIF data from Table 2 indicates a severe multicollinearity problem.

IV. Ridge Regression

Ordinary least squares methods give unbiased estimates and have the minimum variance of all linear unbiased estimators. However, there is no upper bound on what the variance could be, and the presence of multicollinearity could produce large variances. Ridge regression is a biased estimation technique used to attain a substantial reduction in variance with an increase in the stability of the coefficients. If the correlation matrix is reduced, then the variance

$$\mathrm{var}(\underline{\mathbf{b}}) = \sigma^2 (\mathbf{X}'\mathbf{X})^{-1} \tag{7}$$

is improved and the stability of the coefficients is increased. Ridge regression uses this idea.

Variables x and y in Eq. (2) must first be standardized (centered), making them dimensionless relative to an average value

$$x^*_{ij} = \frac{x_{ij} - \bar{x}_j}{\sqrt{\sum_{i=1}^{n} (x_{ij} - \bar{x})^2}} \tag{8}$$

$$y^*_{ij} = \frac{y_i - \bar{y}}{\sqrt{\sum_{i=1}^{n} (y_i - \bar{y})^2}} \tag{9}$$

where i is the number of points ($i = 1, 2, \ldots, n$) and j is the number of parameters ($j = 1, 2, \ldots, k$). The new standardized model becomes

$$\underline{\mathbf{y}}^* = \mathbf{X}^* \underline{\beta}^* + \underline{\epsilon} \tag{10}$$

and the solution for the least squares estimate $\underline{\mathbf{b}}^*_{LS}$ is

$$\underline{\mathbf{b}}^*_{LS} = (\mathbf{X}^{*\prime}\mathbf{X}^*)^{-1} \mathbf{X}^* \underline{\mathbf{y}}^* \tag{11}$$

where $\mathbf{X}^{*\prime}\mathbf{X}^*$ is the correlation matrix, as stated previously.

The ideal correlation matrix is the identity matrix, $\mathbf{I}$. If multicollinearity exists, high correlation values exist so the diagonal elements do not dominate and there are large off-diagonal values. To make the correlation matrix values approach the identity matrix, the ridge estimator is introduced:

$$\underline{\mathbf{b}}^*_R = (\mathbf{X}^{*\prime}\mathbf{X}^* + k\mathbf{I})^{-1} \mathbf{X}^* \underline{\mathbf{y}} \tag{12}$$

where $\mathbf{I}$ is the identity matrix and k is a value greater than or equal to zero and is chosen by the user. The term $k\mathbf{I}$ adds a positive constant to the diagonal elements of the correlation matrix in order to make the diagonal elements dominate. Accordingly, the inverse $(\mathbf{X}^{*\prime}\mathbf{X}^* + k\mathbf{I})^{-1}$ will have smaller elements, alleviating past difficulties created by having large elements on the diagonals of the inverse, like large variances. The term k is often referred to as a "shrinkage parameter" since it "shrinks" the effects of the off-diagonal elements. The ridge estimator $\underline{\mathbf{b}}^*_R$ equals the least squares estimator $\underline{\mathbf{b}}^*_{LS}$ when $k = 0$. It can also be easily converted back to $\underline{\mathbf{b}}_R$ (dimensioned) by a simple transformation.

Ridge regression is called a biased estimation technique since the ridge estimators $\underline{\mathbf{b}}^*_R$ are biased. Proper selection of the shrinkage parameter minimizes the negative effect of large bias while maintaining a ridge estimator variance that is significantly less than the least squares estimator. As the shrinkage parameter increases, the bias of the ridge estimator increases and its variance decreases.

A subjective method exists for choosing the shrinkage parameter: the ridge trace. Many different values of k are used to compute $\underline{\mathbf{b}}^*_R\ (k)$, and then each $\underline{\mathbf{b}}^*_{jR}\ (k)$ is plotted versus k. The more unstable the variable is, the faster it drops off and stabilizes. Gradual changes of the variables over k denote stability. The shrinkage parameter k is chosen so that the estimates are stable. As a rule, the smallest value of k where stability of the coefficients first appears is selected [4], [6].

V. Two Case Studies

Two applications of the ridge regression technique on the systematic error correction model were done using Voyager 1 conical scanning (conscan) offset data. The results were compared to fits obtained using an ordinary linear least squares method. The selected parameters for the linear least squares fit were (refer to Table 1) $P_1, P_7, P_8, P_{12}, P_{13}, P_{14}$, and P_{16}. The parameters selected for the ridge regression cases were $P_8, P_{12}, P_{13}, P_{14}$, and P_{16}. Parameters P_1 and P_7 represent constant cross-elevation and elevation offsets, respectively. In the ridge regression process, these two terms were created by determining the cross-elevation and elevation offset biases.

The first data set uses conscan offset data collected on the 105th day of 1987. As demonstrated in Table 2, this data exhibits a high degree of multicollinearity and would probably benefit from the use of ridge regression. Parameters determined using the linear least squares method are listed in column 1 of Table 3. These parameters exhibit the characteristics associated with multicollinearity, one of them consisting of coefficients that are too large in magnitude (they are too large to be realistic or practical). Shrinkage parameters were selected in 0.005 increments and ranged from 0 to 0.10. Figure 1 shows the use of the ridge trace for the "best" subjective selection of ridge estimators. Stability seems to be reached at approximately $k = 0.02$. The parameters for this shrinkage parameter are listed in column 2 of Table 3. The coefficients have diminished in value, approaching a more realistic representation. Figure 2 compares the residual fit errors obtained in both the linear least squares method and ridge regression. The residual errors are defined in Eq. (5) as the difference between the actual and the fitted pointing offsets. The signatures for both sets of residual errors are similar, indicating incompleteness in the model itself, but the average residual offset for the ridge regression case is nearly zero, and the standard deviations are similar (approximately 0.9 mdeg).

The above example demonstrated how ridge regression can be used to obtain more realistic parameters and fewer overall fitting errors (average error approaching zero). Multicollinearity also causes the parameters to be unstable. Conscan offset data collected from Voyager 1 tracks on the 105th and 106th days should yield similar results. No changes were made to any part of the antenna mechanical subsystem between these two consecutive tracking sessions (for example, the same *a priori* systematic error correction table and autocollimators were employed in both cases), yet the parameters determined using the linear least squares fitting method (listed in columns 1 and 3 of Table 3) seem to indicate otherwise. The parameters not only differ in sign, but also differ radically in magnitude. Parameters determined using ridge regression (columns 2 and 4 of Table 3) are in closer agreement in both magnitude (off by a small factor–3 or 4–rather than 10 or 20) and sign, and also yield similar overall fits (same average and standard deviation).

VI. Conclusion

The ridge regression technique was shown to be useful in minimizing the effects of multicollinearity. For the two examples given, it generated stable coefficients for similar sets of data, provided coefficients that were more realistic in magnitude, and gave an overall fit with average residual errors near zero. Although these are good results in terms of coefficient characteristics, the overall fitting results using ridge regression were no better than the linear least squares results since the signatures resulting from the two methods exhibited analogous trends. A technique such as variable selection or prediction may be needed in order to get a more optimal model and a better parameter selection procedure. In any case, the problem of multicollinearity must still be addressed and resolved.

References

[1] P. Stumpff, "Astronomical Pointing Theory for Radio Telescopes," *Klein-Heibacher Berichte*, vol. 15, Fornmoldeteshnischon Zentralamt, Darmstadt, West Germany, pp. 431–437, 1972.

[2] M. L. Meeks, H. A. Ball, and A. B. Hull, "The Pointing Calibration of the Haystack Antenna," *IEEE Transactions on Antennas and Propagation*, vol. AP-16, no. 6, pp. 746–751, November 1968.

[3] C. N. Guiar, F. L. Lansing, and R. Riggs, "Antenna Pointing Systematic Error Model Derivations," *TDA Progress Report 42-88*, vol. October–December 1986, Jet Propulsion Laboratory, Pasadena, California, pp. 36–46, February 15, 1987.

[4] R. B. Myers, *Classical and Modern Regression with Applications*, Boston: Duxbury, 1986.

[5] E. Kreysig, *Advanced Engineering Mathematics*, New York: John Wiley and Sons, 1983.

[6] A. J. Bush, "Ridge: A Program to Perform Ridge Regression Analysis," *Behavior Research Methods and Instrumentation*, vol. 12, no. 1, pp. 73–74, 1980.

Table 1. Systematic pointing error sources and model terms

Error source	Model function	
	Cross-elevation error	Elevation error
Az collimation	P_1	–
Az encoder fixed offset	P_2 cos (el)	–
Az/el skew	P_3 sin (el)	–
Az axis tilt	P_4 sin (el) cos (az)	$-P_4$ sin (az)
Az axis tilt	P_5 sin (el) sin (az)	P_5 cos (az)
El encoder fixed offset	–	P_7
Gravitational flexure	–	P_8 cos (el)
Residual refraction	–	P_9 cot (el)
Az encoder scale error	P_{10} (az/360) cos (el)	–
	Cross-declination error	Declination error
HA/dec axis skew	$-P_{11}$ sin (dec)	–
HA axis tilt	P_{12} sin (HA) sin (dec)	P_{12} cos (HA)
HA axis tilt	$-P_{13}$ cos (HA) sin (dec)	P_{13} sin (HA)
HA feed offset	$-P_{14}$	–
Gravitational flexure	P_{15} cos (p) cos (el)	$-P_{15}$ sin (p) cos (el)
Declination feed offset	–	P_{16}
Gravitational flexure	P_{17} sin (p) cos (el)	–
Gravitational flexure	–	$-P_{18}$ cos (p) cos (el)
Gravitational flexure	$-P_{19}$ sin (el)	–
Gravitational flexure	–	P_{20} sin (el)
HA encoder bias	P_{21} cos (dec)	–

Note: (1) Uppercase P refers to parameter value; lowercase p refers to paralectic angle.
(2) Az = azimuth angle; el = elevation angle; dec = declination angle; HA = hour angle.

Table 2. Sample correlation matrix and variance inflation factors (for Voyager 1 conscan offset data from 105th day of 1987)

	Correlation matrix					VIF
Variable	8	12	13	14	16	
8	1.0000	0.8653	−0.9911	−0.9954	0.9937	747.6
12	0.8653	1.0000	−0.8624	−0.8690	0.8994	103.9
13	−0.9911	−0.8624	1.0000	0.9981	−0.9959	2783.1
14	−0.9954	−0.8690	0.9981	1.0000	−0.9966	929.0
16	0.9937	0.8994	−0.9959	−0.9966	1.0000	4377.6

Table 3. Model parameters for two Voyager 1 conscan offset data sets (105th and 106th days of 1987) using linear least squares and ridge regression (units are in millidegrees)

Parameter (P)	Day 105		Day 106	
	Linear least squares (1)	Ridge regression (2)	Linear least squares (3)	Ridge regression (4)
1	−451.93	24.08*	−0.55	18.68*
7	−141.18	29.58*	17.86	17.13*
8	−271.78	−13.45	−27.22	−29.11
12	−114.05	4.28	1.40	1.34
13	240.03	−4.10	−0.65	−17.24
14	−557.56	4.85	6.19	15.26
16	103.29	0.25	0.49	0.89

*Ridge regression parameters P_1 and P_7 are created by determining the cross-elevation and elevation biases.

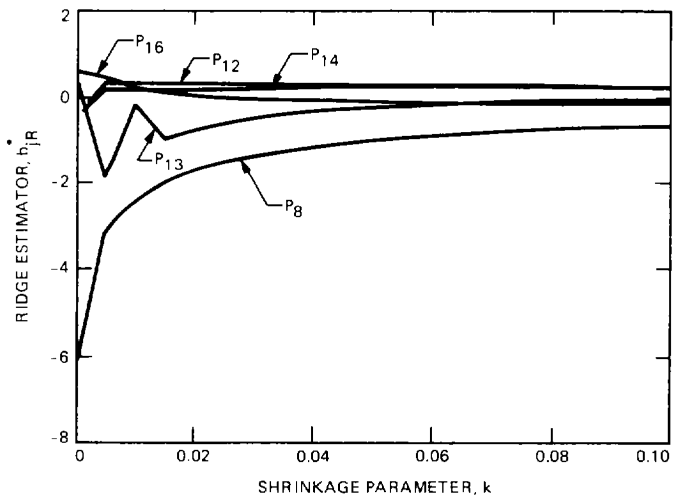

Fig. 1. Ridge trace for conscan offset data from a Voyager 1 track collected on 105th day of 1987 shows parameters reaching stability at approximately k = 0.02

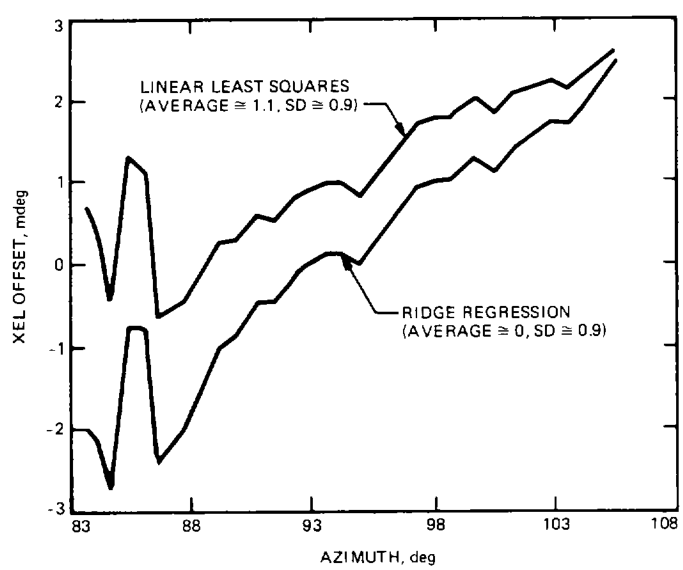

Fig. 2. Comparison of residual pointing errors (difference between actual and fitted offsets) using the linear least squares and ridge regression for data collected on 105th day of 1987